

# EXHIBIT B

**EXHIBIT B**  
**DOCUMENT PRODUCTION FORMAT**

PRODUCTION FORMAT

- 1) All documents will be produced with bates numbers that shall be unique for each page. Bates numbers shall contain a prefix that can be readily attributed to the producing party. Based numbering should be sequential. If a bates number is skipped, the producing party will provide information about the missing bates numbers.
- 2) All documents produced from ***hard copy materials*** should be provided in the following format:
  - a) **Unique ID.** Each image and OCR text file should have a unique file name. Each image will be assigned a unique file name based on the production bates number assigned to the page. For OCR text files, the unique file name will be the bates number assigned to the first page of the document.
  - b) **Image Format.** All documents should be scanned and produced in single-page Group IV TIF format. Images should be at least 300 dpi. Documents containing meaningful color should be produced in color image format (JPG).
  - c) **OCR Text.** For each document produced, multi-page OCR with page breaks should be provided as a separate text file. The filename of the OCR text file should match the bates number assigned to the first page of the document's correlating TIF file (e.g., ABC0000001.TIF would correlate to ABC0000001.txt). The text files will not contain the redacted portions of documents.
  - d) **Unitizing Documents.** Paper documents should be logically unitized. Distinct documents should not be merged into a single record, and single documents should not be split into multiple records. The parties will make their best efforts to have their vendors unitize documents correctly prior to OCR and will commit to address situations where there are improperly unitized documents.
  - e) **Objective Coding Fields.** The following objective coding fields should be provided to the extent practicable: (1) beginning bates number; (2) ending bates number; (3) source/custodian; and (4) page count.
  - f) **Database Load Files / Cross-Reference Files.** Documents should be provided with the load files described in paragraph 3(f) below.
- 3) All documents produced from ***electronic data*** should be provided in the following format:
  - a) **Unique ID.** Each image, native file and extracted text file should have a unique file name. Each image will be assigned a unique file name based on the production bates number assigned to the page. For native files and extracted text files, the unique file name will be the bates number assigned to the first page of the document.
  - b) **Image Format.** All documents should be produced in single-page Group IV TIF format. Images should be at least 300 dpi. Documents containing meaningful color

should be produced in color image format (JPG). The producing party shall have the option to produce Excel spreadsheets in native format.

- c) **Native Format.** All parties reserve the right to request, within reason, select electronic documents in native format. All requests must be made in writing and contain the bates number(s) assigned to the document(s) along with the file name as contained in the metadata. Excel spreadsheets and PowerPoint presentations shall automatically be produced in native form within each production, with a placeholder TIF image that includes: (1) "This document produced in native format" and (2) a Bates number. If, however, a document that would otherwise be produced in native format is redacted, it shall be produced in TIF format, and OCR text shall be produced from the redacted image(s).
- d) **Extracted Text.** For each document produced, the extracted text of the document should be provided. The text should be extracted directly from the native file (except for redacted documents) and be provided in multi-page format so that a single text file correlates to a single document. In the case of redacted documents, the text files will not contain the redacted portions of documents, and the text will be obtained by OCR from the redacted TIF image. Any redacted, privileged material should be clearly labeled to show the redaction on the TIF image.
- e) **Metadata Fields.** The metadata fields identified in Exhibit "A" should be provided if they exist. The parties will make their best efforts to have their vendors provide the requested metadata field information.
- f) **Database Load Files / Cross-Reference Files.** There should be three (3) load files included with the production:
  - i) For productions to Shionogi: An ASCII text file containing the field data for loading into any litigation support database. The field data should include the following delimiters:

Field Separator		(Piping)
Quote Character	^	(Caret)
Multi-Entry Delimiter	;	(Semi-Colon)
<Return> Value in Data	~	(Tilde)

For productions to InterMune: A Concordance DAT load file using the Concordance default delimiters, as follows:

"□" (ANSI 020) for comma  
 "p" (ANSI 254) for quote  
 "®" (ANSI 174) for newline  
 MM/DD/YYYY for Dates

In either case, the data load file should include all of the metadata fields referenced in paragraph 2(e) above. The first line of the data load file should contain field information. The data load file should correlate to the media volume name (e.g., ABC\_0001.txt or ABC\_0001.dat).

There should not be any gaps in the bates numbering for all documents contained in the data load file unless those gaps are identified in an explanatory letter/email or a privilege log provided by counsel.

- ii) The image load file should be in both OPT and LFP formats. The files should contain the full path and related control number for each image as follows:

IM,[control number],[D (if new document boundary; leave blank if no boundary demarcation is needed)],0,@[media volume name];[first level directory]\[second level directory]\[third level directory];[tiff file name];2

```
IM,ABC0000001,D,0,@ABC_0001;Images\00\00;ABC0000001.tif;2
IM,ABC0000002,A,0,@ABC_0001;Images\00\00;ABC0000002.tif;2
IM,ABC0000003, ,0,@ABC_0001;Images\00\00;ABC0000003.tif;2
IM,ABC0000004,D,0,@ABC_0001;Images\00\00\ABC0000004.jpg;4
IM,ABC0000005, ,0,@ABC_0001;Images\00\00\ABC0000005.jpg;4
IM,ABC0000006,A,0,@ABC_0001;Images\00\00;ABC0000006.tif;2
```

Final document unitization should be reflected in the LFP file. Default boundary demarcation should be:

D = Document  
A = Attachment

The Opticon (OPT) load file is a page level load file, with each line representing one image. Below is a sample:

```
REL00001,REL01,D:\IMAGES\001\REL00001.TIF,Y,,,3
REL00002,REL01,D:\IMAGES\001\REL00002.TIF,,,,
REL00003,REL01,D:\IMAGES\001\REL00003.TIF,,,,
REL00004,REL01,D:\IMAGES\001\REL00004.TIF,Y,,,2
REL00005,REL01,D:\IMAGES\001\REL00005.TIF,,,,
```

The fields are, from left to right:

Field One – (REL00001) – The page identifier  
Field Two – (REL01) – The volume identifier – not required  
Field Three – (D:\IMAGES\001\REL00001.TIF) – a path to the image to be loaded  
Field Four – (Y) – Document marker – a “Y” indicates the start of a unique document.  
Field Five – (blank) – unused

The files should be included on the media volume and have the same volume name (e.g., media volume ABC\_0001 will contain ABC\_0001.lfp and ABC\_0001.opt).

Every TIF or JPG must be referenced in the production’s corresponding image load file.

- iii) The load file for extracted text or OCR should be in ASCII format and have an LST extension (“Control List”). The Control List consists of the image key for a document, followed by the path to the full text document. The image key reference contained within the Control List must match the BegDoc# field data as well as the image key assigned in the image load file. An example of the syntax for the Control List is as follows:

ABC0000001,E:\Images\00\00\ABC0000001.txt  
ABC0000002,E:\Images\00\00\ABC0000002.txt  
ABC0000004,E:\Images\00\00\ABC0000004.txt  
ABC0000006,E:\Images\00\00\ABC0000006.txt

The Control List file name should correlate to the media volume (e.g.,  
ABC\_0001.LST).

**EXHIBIT "A" – METADATA FIELDS TO INCLUDE FOR ELECTRONIC SOURCE DATA**

	<i>Field Name</i>	<i>Field Type</i>	<i>Description</i>
1	BegDoc#	Note	The unique identification number associated with the first page of a document.
2	EndDoc#	Note	The unique identification number associated with the last page of a document.
3	BegAttch	Note	The unique identification number associated with the first page of the <u>range</u> that includes all family members.
4	EndAttch	Note	The unique identification number associated with the last page of the <u>range</u> that includes all family members.
5	PC	Note	Identifies the documents as a Parent (P) or a Child (C)
6	Family Range	Note	The unique identification numbers associated with the first page and the last page of the range that includes all family members.
7	PgCount	Integer	The number of pages contained within the document.
8	DocDate	Date	The date of the document (GMT). [ <b>NOTE:</b> For electronic files, parties should determine which date is most accurate within their processing software – usually date last modified or date sent.]
9	Author/From	Multi-Entry	The display name of the document's author.
10	Recip/To	Multi-Entry	The display name of the recipient(s) of a document.
11	CC	Multi-Entry	The display name of copyee(s) of a document.
12	BCC	Multi-Entry	The display name of blind copyee(s) of a document.
13	Subject	Note	The subject of a document usually contained in the "subject" line of the email.
14	Custodian	Note	The custodian or source of a document.
15	DocLink	Note	Document Link field. This field will include the path to the native file of the document on the production media. If the native file is not produced, this field will be blank.
16	FileExt	Note	The file extension of a document.
17	FileName	Note	The file name of a document.
18	FileType	Note	The software application needed to open the native file (e.g., Word, Excel, etc.).
19	MD5Hash	Note	The unique value assigned to a document by the vendor software system.
20	TextPath	Note	Full text path field containing the relative path of the associated TXT file.
21	SourcePath	Note	The full folder file path to the document's original file.
22	Primary Language	Note	This field will include the language in which a majority of the document is written (e.g., English, Japanese, French, etc.)
23	Languages	Multi-Entry	This field will include all the languages in which the document is written.